

**Bibliograficzne bazy danych i ich rola w rozwoju nauki****II Konferencja naukowa Konsorcjum BazTech****Poznań, 17-19 kwietnia 2013**

Katarzyna Zamłyńska
Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
Uniwersytet Warszawski

DML-PL Polska Matematyczna Biblioteka Cyfrowa

Katarzyna Zamłyńska pracuje w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego od 2003 r. Zajmuje się bibliotekami cyfrowymi, m.in. Polską Matematyczną Biblioteką Cyfrową. Analitik systemu POL-index.

Streszczenie: DML-PL jest istniejącą od ponad 10 lat otwartą biblioteką polskich publikacji matematycznych, mającą swoich użytkowników na całym świecie. Udostępnia pełne teksty artykułów i książek – zarówno tych najnowszych, jak i archiwalnych. DML-PL jest częścią nowo powstałej Europejskiej Matematycznej Biblioteki Cyfrowej EuDML, która zapewnia dostęp do europejskiej spuścizny matematycznej. W artykule opisano zawartość biblioteki DML-PL oraz statystyki jej użytkowania.

Słowa kluczowe: biblioteki cyfrowe, otwarty dostęp, publikacje matematyczne, repozytoria wiedzy, YADDA

Abstract: DML-PL has existed for over 10 years, it is a widely used open digital library of Polish mathematical publications. It provides access to full texts of journal articles and books, archives and current issues. DML-PL is a part of the new open European Digital Mathematics Library – EuDML, which provides access to the European heritage of mathematics literature. In this article we describe the content of the DML-PL library and some user statistics.

Keywords: digital libraries, open access, mathematics publications, repositories of knowledge, YADDA

Prezentacja

Wstęp

Prace nad utworzeniem Polskiej Matematycznej Biblioteki Cyfrowej – DML-PL (<http://pldml.icm.edu.pl>¹) rozpoczęły się ponad 10 lat temu. Celem projektu było zamieszczenie na stronach biblioteki, na zasadach otwartego dostępu, wydanych w Polsce publikacji stricte matematycznych i związanych z matematyką.

Początkowo prace nad biblioteką polegały głównie na skanowaniu archiwalnych tomów polskich czasopism matematycznych i udostępnianiu ich w Internecie. W tym okresie gromadzono jedynie podstawowe metadane artykułów: tytuł, numery stron, numery tomu i zeszytu, rok wydania oraz imiona i nazwiska autorów. Rdzeniem DML-PL przez wiele lat była bardzo prosta baza MySQL. Biblioteka umożliwiała wtedy jedynie przeglądanie czasopism oraz proste wyszukiwanie po nazwiskach autorów i słowach w tytule artykułu. Mimo skromnych możliwości DML-PL zaczęła cieszyć się coraz większą popularnością².

¹ Wszystkie odniesienia do stron internetowych przedstawiają wersję aktualną w dn. 17.06.2013 r.

² ZAMŁYŃSKA, K., TARKOWSKI, A., ROSIEK, T. Evolution of Mathematical Collection Polish Virtual



W kolejnych etapach prac nad DML-PL znacznie rozszerzono jej funkcjonalność. Od maja 2012 r. biblioteka w całości dostępna jest na platformie YADDA³. Około 10% artykułów uzupełniono już o pełne metadane, w tym abstrakty, bibliografie i słowa kluczowe. Dzięki migracji na platformę YADDA, DML-PL spełnia obecnie wymogi nowoczesnej biblioteki cyfrowej, a artykuły prezentowane są w niej w sposób znacznie wygodniejszy dla użytkowników.

EuDML

W 2008 r. rozpoczęły się prace związane z ogólnoeuropejskim projektem pod nazwą „EuDML — Europejska Cyfrowa Biblioteka Matematyczna” („European Digital Mathematics Library” — <http://eudml.org>). Jednym z jego głównych partnerów było Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego. Inicjatywa miała na celu zgromadzenie zasobów wielu istniejących krajowych wirtualnych bibliotek matematycznych w jednym repozytorium⁴. W projekcie brało udział 14 europejskich instytucji, których lista zamieszczona została w tabeli 1.

Projekt zakończył się w roku 2013. Powstałe repozytorium również opiera się na platformie YADDA. Obecnie Europejska Cyfrowa Biblioteka Matematyczna udostępnia ponad 235 000 artykułów i rozdziałów z książek, z czego 98% na zasadach otwartego dostępu. Na stronach EuDML udostępnionych zostało również ponad 2,9 mln rekordów pochodzących z jednej z największych matematycznych baz bibliograficznych — Zentralblatt.

Tab. 1. Lista partnerów w projekcie EuDML

1. Université Joseph-Fourier, Francja	2. Édition Diffusion Presse Sciences, Francja
3. FIZ: Fachinformationszentrum /Zentralblatt MATH, Niemcy	4. Universidade de Santiago de Compostela, Hiszpania
5. Instituto Superior Tecnico, Portugalia	6. BAS, Bułgaria
7. University of Birmingham, Wielka Brytania	8. Matematický Ústav, AV CR, Czechy
9. Masarykova Univerzita, Czechy	10. Ionian University, Grecja
11. ICM, Uniwersytet Warszawski, Polska	12. Made Media Ltd, Wielka Brytania

Library. *Mathematics in Computer Science* 2010, Vol. 3, nr 3, s. 265–278.

³ ZAMŁYŃSKA, K., BOLIKOWSKI, Ł., ROSIEK, T. Migration of the Mathematical Collection of Polish Virtual Library of Science to the YADDA platform. W: *DML 2008 Towards Digital Mathematics Library, Birmingham, UK, July 27th, 2008 Proceedings*. Brno, 2008, s. 127–130.

⁴ SYLWESTRZAK i in. EuDML — Towards the European Digital Mathematics Library. W: SOJKA, P. *Proceedings of DML 2010*. Brno, 2010, s. 11–24.



Bibliograficzne bazy danych i ich rola w rozwoju nauki

II Konferencja naukowa Konsorcjum BazTech

Poznań, 17-19 kwietnia 2013



13. Consejo Superior de Investigaciones Científicas, Hiszpania	14. Centre National de la Recherche Scientifique, Francja
--	---

Źródło: Opracowanie własne.

Od kwietnia 2012 r. zasoby DML-PL dostępne są również za pośrednictwem EuDML i aktualizowane na bieżąco. Dzięki temu każde nowo wprowadzone do polskiej biblioteki czasopismo i każdy nowo dodany tom jest automatycznie widoczny w EuDML.

Zaletą Europejskiej Cyfrowej Biblioteki Matematycznej jest również zaawansowane oprogramowanie, dzięki czemu korzystanie z jej zasobów staje się niezwykle proste. Jedną z przydatnych funkcji jest możliwość zaawansowanego wyszukiwania wzorów oraz czytelnego prezentowania formuł matematycznych w tytułach, abstraktach i cytowaniach⁵. Większość artykułów dostępnych w postaci skanów została poddana procedurze OCR, dzięki czemu możliwie jest ich pełnotekstowe przeszukiwanie. Biblioteka wykorzystuje również mechanizm rozwiązywania cytowań i linkowania pozycji bibliograficznych do rekordów artykułów znajdujących się w EuDML lub w bazie Zentralblatt.

Zawartość DML-PL

W Polskiej Matematycznej Bibliotece Cyfrowej dostępnych jest obecnie ponad 15 000 artykułów i 68 książek. Większą część zbiorów DML-PL stanowią archiwalne numery czasopism matematycznych wydawanych aktualnie przez Instytut Matematyczny Polskiej Akademii Nauk. Najważniejsze z nich to:

- „Fundamenta Mathematicae”,
- „Studia Mathematicae”,
- „Acta Arithmetica”,
- „Colloquium Mathematicum”,
- „Applicationes Mathematicae”,
- „Annales Polonici Mathematici”,
- „Banach Center Publications”.

Pierwsze trzy czasopisma z powyższej listy zostały założone w okresie międzywojennym przez matematyków związanych z Polską Szkołą Matematyczną. Publikowane są w nich nie tylko artykuły naukowe, lecz także nekrologi, sprawozdania i kroniki, stąd stanowią one interesujący materiał nie tylko dla matematyków, ale i historyków.

⁵ BARBINHA, J. i in. Project EuDML – A First Year Demonstration. W: *Intelligent Computer Mathematics*. Heidelberg: Springer 2011, s. 281–284. (Lecture Notes in Computer Science, Vol. 6824).



Bibliograficzne bazy danych i ich rola w rozwoju nauki

II Konferencja naukowa Konsorcjum BazTech

Poznań, 17-19 kwietnia 2013



Najstarsze czasopismo dostępne w DML-PL to „Prace Matematyczno-Fizyczne”. Zostało ono założone w 1888 r. i w owym czasie wydawane było przez Samuela Dicksteina. Czasopismo przestało ukazywać się dopiero w roku 1952.

Jedynie dwa czasopisma udostępniają na stronach biblioteki swoje numery bieżące. Są to „Control and Cybernetics” wydawane przez Instytut Badań Systemowych Polskiej Akademii Nauk oraz „International Journal of Applied Mathematics and Computer Sciences”, którego wydawcą jest Uniwersytet Zielonogórski. Stanowią one około 5% zbiorów biblioteki, istnieją jednak plany mające na celu włączenie kolejnych bieżących wydawnictw matematycznych.

Okolo 70% zbiorów biblioteki stanowią skany opatrzone bardzo ubogimi metadanymi. Pozostałe numery archiwalne stanowią pliki PDF skompilowane z oryginalnych plików w formacie TeX lub LaTeX, w przypadku których istnieje możliwość łatwego wygenerowania metadanych. Okolo 10% zasobów opatrzone pełnym opisem bibliograficznym.

W bibliotece cyfrowej dostępne są również dwie serie książek: *Monografie Matematyczne* oraz *Dissertationes Mathematicae*, która wcześniej ukazywała się pod polskim tytułem *Rozprawy Matematyczne*. Obie serie wydawane są obecnie przez Instytut Matematyczny Polskiej Akademii Nauk. W zasobach są m.in. dwie bardzo ważne dla matematyków książki: drugi tom *Dzieł zebranych* Stefana Banacha oraz trzy tomy *Pism* Mariana Smoluchowskiego.

Statystyki

W ciągu pierwszych pięciu lat istnienia DML-PL udostępnionych zostało 85% jej aktualnych zasobów. W późniejszych latach nastąpiło wyraźne spowolnienie tempa prac. Liczbę artykułów dostępnych w bibliotece w kolejnych latach przedstawia rysunek 1⁶. W najgorszym okresie (rok 2009) udostępnione zostały tylko trzy nowe tomy, z powodu braku środków na skanowanie numerów archiwalnych oraz koncentracji na pracach związanych ze wzbogacaniem metadanych publikacji i przeniesieniem biblioteki na platformę YADDA.

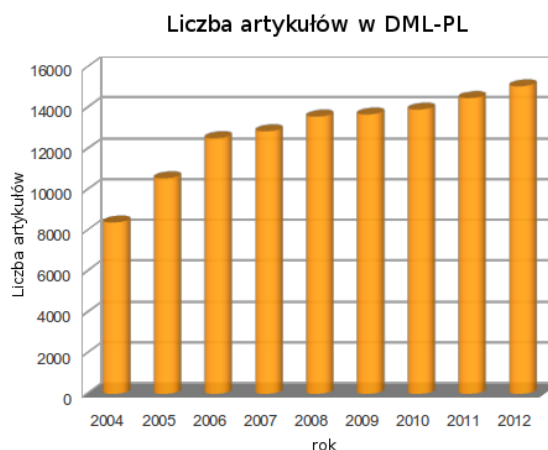
⁶ Zachowane zostały dane dotyczące wejść na stronę od października 2004 r. Wcześniejsze dane zostały utracone. Wszystkie statystyki powstały w okresie od października 2004 r. do grudnia 2012 r.



Bibliograficzne bazy danych i ich rola w rozwoju nauki

II Konferencja naukowa Konsorcjum BazTech

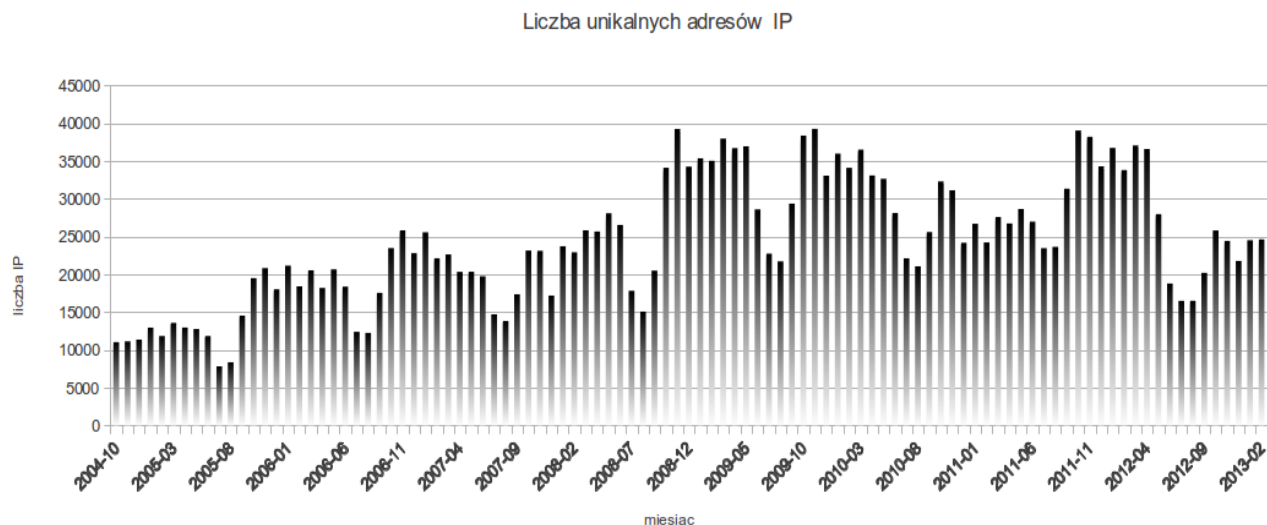
Poznań, 17-19 kwietnia 2013



Rys. 1. Liczba artykułów dostępnych w DML-PL w latach 2004-2012

Źródło: Opracowanie własne.

Mimo tego, że w bibliotece nie pojawiały się nowe pozycje, liczba jej użytkowników nie malała. Rysunek 2 przedstawia liczbę unikalnych numerów IP użytkowników łączących się ze stroną DML-PL w kolejnych miesiącach jej funkcjonowania.

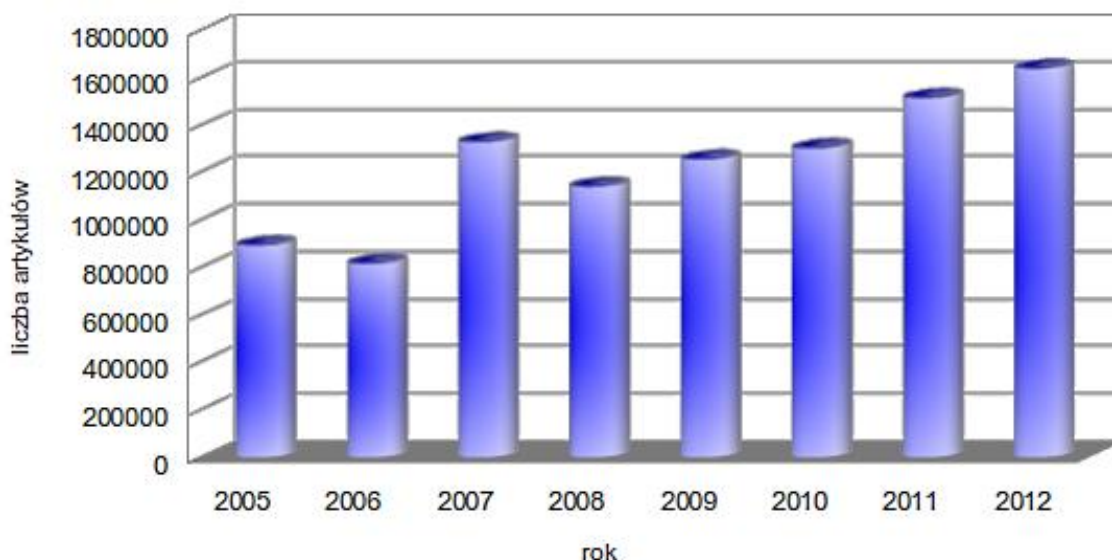


Rys. 2. Liczba połączeń z unikalnych numerów IP w zależności od miesiąca

Źródło: Opracowanie własne.

W ciągu miesiąca rejestrowanych jest średnio 24 000 połączeń z unikalnym numerem IP, średnia roczna to ponad 235 000 numerów IP. Od początku prowadzenia statystyk, czyli od października 2004 r., z biblioteki skorzystało ponad 1,6 mln unikalnych użytkowników.

Na rysunku 2 zaobserwować można, że liczba użytkowników DML-PL nieznacznie maleje od kwietnia 2012 r. Jest to prawdopodobnie spowodowane udostępnieniem zasobów biblioteki na stronach EuDML, gdzie przechowywane są jedynie rekordy bibliograficzne artykułów, zaś pełne teksty ściągane są bezpośrednio z serwerów poszczególnych bibliotek. Spadek liczby użytkowników DML-PL nie oznacza więc spadku popularności polskich czasopism matematycznych. Potwierdza to rysunek 3, na którym widać, że liczba pobranych z serwerów DML-PL plików PDF w 2012 r. wzrosła.



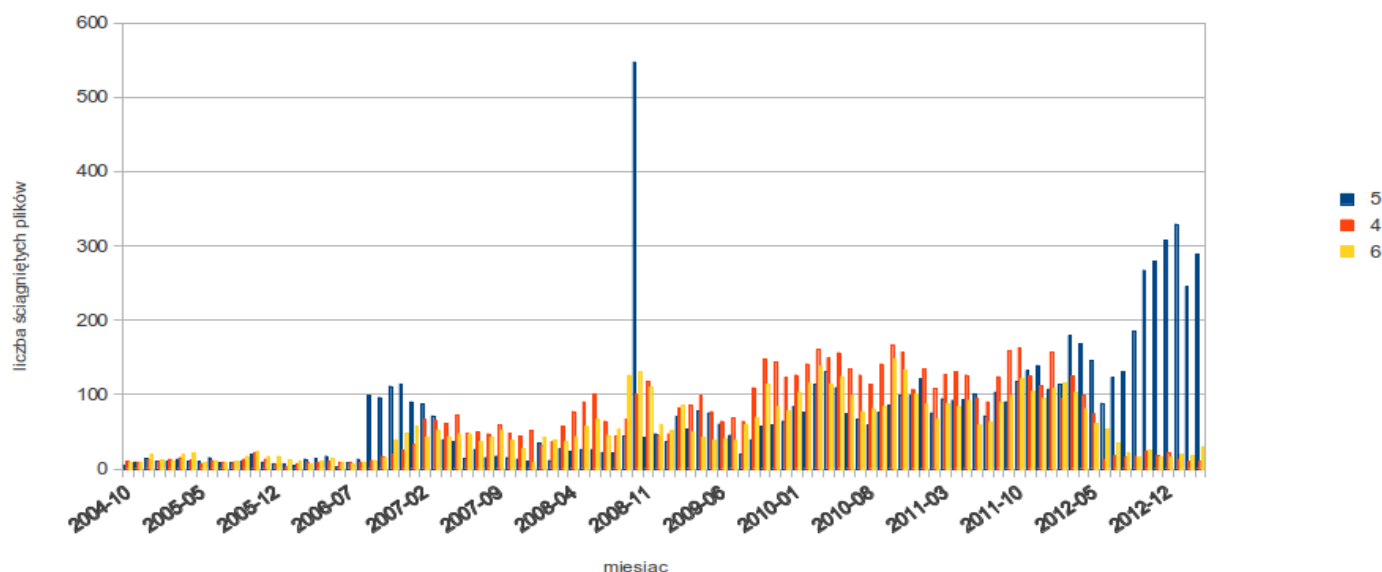
Rys. 3. Liczba pobranych artykułów w formie plików PDF w latach 2005-2012

Źródło: Opracowanie własne.

Zalety platformy YADDA w praktyce

Przez wiele lat biblioteka DML-PL dostępna była za pośrednictwem strony internetowej i funkcjonowała w oparciu o bardzo prostą bazę danych. Oferowała ograniczone możliwości wyszukiwania: większość użytkowników przeglądała jej zasoby, aby dotrzeć do konkretnej pracy. Analiza pobrań plików pokazuje, jak często użytkownicy biblioteki mylili się oraz jak przeniesienie biblioteki na platformę YADDA poprawiło tę sytuację.

Na rysunku 4 przedstawiona została miesięczna liczba pobrań trzech artykułów z czasopisma „International Journal of Applied Mathematics and Computer Sciences” w latach 2004–2012. Kolorem niebieskim oznaczony jest artykuł autorstwa E. Delaleau, J. Louisa i R. Ortegi *Modeling and control of induction motors*. Artykuł ten zamieszczony był jako piąty na liście artykułów dostępnych w bibliotece. Kolorem czerwonym i żółtym oznaczone są dwa sąsiednie artykuły: czwarty i szósty.



Rys. 4. Liczba pobrań trzech artykułów z czasopisma *International Journal of Applied Mathematics and Computer Science* 11, s. 77–104, 105–129, 131–164
Źródło: Opracowanie własne.

Publikacja E. Delaleau należy do najczęściej ściąganych artykułów w DML-PL. Nie wiadomo wprawdzie, czym spowodowany był widoczny wzrost liczby pobrań pliku w końcu 2008 r., artykuł ten był jednak pobierany relatywnie często nawet z pominięciem tego okresu. Przez kilka lat liczba pobrań artykułów występujących na stronie bezpośrednio przed i po tym artykule, a więc czwartego i szóstego artykułu na liście, była równie duża, a nawet większa niż artykułu piątego.

W starej wersji biblioteki kliknięcie w tytuł powodowało ściągnięcie pliku PDF publikacji. Popularność artykułu czwartego i szóstego spadła tymczasem w maju 2012 r., czyli po przeniesieniu biblioteki na platformę YADDA. Wynika z tego, że użytkownicy biblioteki znacznie łatwiej obecnie trafiają na poszukiwany artykuł.

Plany na przyszłość

Po wielu latach starań, w roku 2013 udało się pozyskać środki na rozbudowę DML-PL. W ciągu najbliższego roku zeskanowane i udostępnione zostaną brakujące numery archiwalne czasopism dostępnych w bibliotece. Dzięki temu powiększy się ona o ponad 400 tomów.

Rozszerzenie Biblioteki Matematycznej będzie również możliwe dzięki współpracy nawiązanej w ramach projektu DML-PL z bazą bibliograficzną BazTech (<http://baztech.icm.edu.pl>). Udział w projekcie EuDML daje z kolei możliwość



Bibliograficzne bazy danych i ich rola w rozwoju nauki

II Konferencja naukowa Konsorcjum BazTech

Poznań, 17-19 kwietnia 2013



poprawiania jakości pełnych tekstów publikacji archiwalnych, które na bieżąco poddawane są procedurze OCR. Dzięki temu w najbliższym czasie uzupełnione zostaną metadane wielu z nich. Możliwe stanie się również pełnotekstowe przeszukiwanie.

Bibliografia:

1. BARBINHA, J. i in. Project EuDML – A First Year Demonstration. W: *Intelligent Computer Mathematics*. Heidelberg: Springer 2011, s. 281–284. (Lecture Notes in Computer Science, Vol. 6824).
2. SYLWESTRZAK, W. i in. EuDML – Towards the European Digital Mathematics Library. W: SOJKA, P. *Proceedings of DML 2010*. Brno: Masaryk University, 2010, s. 11–24.
3. ZAMŁYŃSKA, K., BOLIKOWSKI, Ł., ROSIEK, T. Migration of the Mathematical Collection of Polish Virtual Library of Science to the YADDA platform. W: *DML 2008 Towards Digital Mathematics Library, Birmingham, UK, July 27th, 2008 Proceedings*. Brno: Masaryk University, 2008, s. 127–130.
4. ZAMŁYŃSKA, K., TARKOWSKI, A., ROSIEK, T. Evolution of Mathematical Collection Polish Virtual Library. *Mathematics in Computer Science* 2010, Vol. 3, nr 3, s. 265–278.

Zamłyńska, K. DML-PL Polska Matematyczna Biblioteka Cyfrowa. W: Bibliograficzne bazy danych i ich rola w rozwoju nauki. II Konferencja naukowa Konsorcjum BazTech, Poznań, 17-19 kwietnia 2013 [online]. Stowarzyszenie EBIB, 2013 [Dostęp: 30.08.2013]. Materiały konferencyjne EBIB, nr 24, Dostępny w World Wide Web: http://open.ebib.pl/ojs/index.php/Mat_konf/article/view/48/. ISBN 978-83-63458-06-5.